Grasslands Division, DSIR, Private Bag, Christchurch, New Zealand.

# DESCRIPTION OF VEGETATION USING VISUAL RANKING OF SPECIES

**Summary:** A vegetation sampling technique is described where the prime measurement is only the listing of the relative order of abundance (rank) of species at a series of sample points and is estimated visually. Such data are appropriate for ordinal methods of statistical analysis. Repeated measurements show that the observer is only a small component of variation in rank values. The analysis of rank data for individual species is shown for deriving summary parameters, deriving association between species, and determining significance of changes in ranking with respect to qualitative or quantitative treatment variables. Vegetation types can be defined by the rank order of constituent species, and a hierarchial procedure is described for subdivision of sample data. Repeated measurements at regular intervals allow the application of the transition matrix approach to the study of vegetation trends.

**Keywords:** vegetation sampling; visual estimates; rank ordinal data; statistics; transition matrix.

## Introduction

Plant ecologists who wish to describe species abundance, vegetation and vegetation trends, are inevitably faced with conflict. There is a need to cover large areas; field time is at a premium; there is frequently initial uncertainty of what are going to be the important features; and it is desirable to keep the sampling broadly based, but with the possibility of suitable quantification once trends are detected. The suggestions developed in this paper follow from the observation that in the company of ecologists conversant with a range of species, naming the most abundant species at a particular site immediately conveys a great deal of information about that site. For instance, to state that the most abundant species is *Festuca novae zelandiae* indicates a high probability that one is referring to a mid-altitude native grassland in the low to moderate rainfall zone, and possibly being used for extensive rangeland agriculture. This indication is further refined as successive species are listed in order of abundance. For example if the second most abundant species is stated to be the adventive *Hieracium praealtum,* then it probably implies a community which has much bare ground an, little-grazing in the past decade. The description and probable dynamics of a vegetation type improves as further species are listed in rank order.

It seemed that this listing of species, and the information conveyed in their ranking, should be amenable to greater analytical development, especially in situations where they are the only type of available or practical observation.

The first section of the paper describes the concept, field technique, the statistical nature of the data produced, and its reliability. The middle section of the paper considers the analysis of the resulting data as it applies to the values for individual species. However the rank values are basically a vegetation characteristic rather than of the constituent species. The last section discusses the definition of vegetation types by species ranking and determining vegetation trends from repeated measurements of sample quadrats using the transition matrix approach.

## Techiques and Data

*Field technique*

Field application involves listing species in order of abundance at a number of sample points within an area. The sampling points may be defined quadrats or be plotless. In one study, described more fully later, fixed 1 m$^2$ quadrats, numbering about 30 per 0.25 ha treatment plot, were used and ranked annually. In another study assessment was made at 5-pace intervals along 100-pace transects of various topographic strata within one region (Scott and Maunsell, 1974). The difference between the two methods is that the number of species may be limited within a quadrat whereas in a plotless sample observations can be extended outwards until the desired number of species is located.

At each point the species are ranked in order of plant abundance - a term that will need to be defined according to the objectives of a particular study. For some it might be the density of each species; for others it might be the contribution to percentage ground cover in terms of protection of the soil surface from erosion. In examples used here the criterion was the visual estimate of above-ground dry matter based on experience from cut quadrats. The method assumes that, whatever criterion is used, visual ranking will give the same rank order as if quantitative measurements were made, e.g., if the species could be cut, dried and weighed. The accuracy of the visual ranking will be discussed further in a later section.

At least five species should be ranked at each sampling point, but the number ranked can be varied according to both the vegetation richness and quadrat size used. In our studies, in modified or partially developed New Zealand tussock grasslands, the first ten most abundant species were ranked and the presence of other species noted. In developed pastures, or in quadrats of less than 1m', it is difficult to get beyond five species. Later sections show that most of the quantitative aspects of the analysis can be obtained by ranking about five species. It is also useful for possible transformation of the rank data to a quantitative scale to have an estimate of the ratio of abundance of any two species in at least a proportion of the samples.

Table 1 illustrates an example of ran kings for five sample points in the field, and shows two characteristics of the method. Firstly, the rank value of a species is only relative to others in the same sample, for example the change in ranking of species 'B' in the second as compared to the other sampling points could relate to either the decrease in abundance of 'B' or the increase in 'C'. Secondly, the number of species ranked will generally be less than the total number of species present in an area so that a particular species will not be ranked on all occasions (for example species 'D' and 'E') and consequently the raw data matrix will generally contain a large number of empty cells.

Table 1: *Example of field notes for visual ranking of the five most abundant species at five sample points. 1 = 1st ranked (most abundant), 2 = 2nd ranked etc. p = species present but not among the five most abundant species (ultimately given a rank of 6). c = species absent at that sampling point*

| Sample | Species | | | | | | | Estimated ratio |
|---|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | G | $5^{th}$ to 1st |
| 1 | 1 | 2 | c | | 3 | 4 | 5 | 0.15 |
| 2 | 1 | 2 | p | 2 | | 5 | 4 | 0.10 |
| 3 | 1 | 2 | 3 | | 4 | | 5 | 0.10 |
| 4 | 1 | 2 | 4 | | p | 3 | 5 | 0.20 |
| 5 | 1 | 2 | | 5 | p | 3 | 4 | 0.05 |

Experience has shown that two types of field error have to be guarded against. The first is species recognition and the consequent 'blind spot' developed by observers - a problem common to any vegetation sampling method. The second is the occasional non-ranking of an abundant species due to observers moving from the ground search of rare species in a previous sample and not standing up, refocusing and starting anew with the larger common species. Data checking indicates a 1-2% occurrence of errors of this latter type.

Before the field rank data can be analysed it is necessary to designate ranks to species present but not abundant, and to species absent in a particular sample. For example, if five ranks were recognised in the field then other species present get transformed to a rank of 6, and those absent to rank 7. Several species may share these latter two ranks.

The use of the field estimation of the ratio of abundance of fifth to first ranked species will be briefly referred to in the discussion.

*Data*

The data used to illustrate the various analysis options came from a study of short tussock *(Festuca novae-zelandiae)* grassland near Twizel in the Mackenzie Country which, at the commencement of an experiment, had been invaded by browntop *(Agrostis capillaris)* and mouse-ear hawkweed *(Hieracium pilosella)*. Trial plots had been set up to compare the effectiveness of vegetation change following sowing with alsike clover *(Trifolium hybridum)* and white clover *(T. repens)*. The treatments were method of clover introduction (strip sowing versus general overdrilling) and season and type of grazing (spring and early summer, late summer and autumn, winter) or rotational stocking. The eight treatments (50 x 50 m) had 30 or more permanent quadrats of 1 x 1 m per plot in which species were ranked each summer from the second to sixth year. The first year was the sowing year with its associated distubance. The site and original vegetation were uniform at a macro-scale but variable at the micro-scale of quadrat size. Hence it was justifiable to regard the quadrats as the sample units even though they were geographically grouped within treatment areas. The data used were a subset of those described more fully elsewhere. In particular, *most* results are only for four common species out of the fifty or so present.

*Reliability of rank observations*

The reliability of the rank data has at least two components: the population variability of the vegetation and hence rank of species at the scale sampled; and the observer consistency in ranking particular quadrats. An indication of the interaction between these two sources of variation was obtained by 3 observers independently ranking species in 50 fixed quadrats on 2 occasions at 3 day intervals.

Table 2: *Sources of variation in field ranking. **= highly significant (p<0. 01).*

a) Change in variance components of species with mean rank

| | a | b | c | d | e | f | g | h |
|---|---|---|---|---|---|---|---|---|
| Mean rank | 1.6 | 1.8 | 2.8 | 5.7 | 6.2 | 7.0 | 8.0 | 9.9 |
| Standard deviation | 0.7 | 1.3 | 1.9 | 3.8 | 4.1 | 3.5 | 3.1 | 3.1 |
| % variation | | | | | | | | |
| - between quadrat ( = vegetation) | 87 | 65 | 68 | 88 | 84 | 77 | 72 | 91 |
| - within quadrat (= other) | 13 | 35 | 32 | 12 | 16 | 23 | 28 | 9 |

b) Variation of standard deviation (SD) in rank

| | Rank | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| SD | 0.6 | 1.5 | 1.5 | 1.5 | 1.9 | 2.0 | 2.4 | 2.3 | 1.9 | 1.8 |

c) Other comparisons of standard deviation

| | Observer | | | Vegetation | |
|---|---|---|---|---|---|
| | I | II | III | Contrasting | Similar |
| Rank 1 | 0.5 | 0.5 | 0.9** | 0.3 | 0.8** |
| Rank 2 | 1.8 | 1.8 | 2.6** | 1.8 | 2.0 ns |

Table 2a gives the mean rank and standard deviation for 8 species of contrasting abundance. The standard deviations of rank were greatest at mid-rank, and of the order of three rank positions. The variability of these mid-ranked species indicates the large number of samples that would be needed to get reliable mean ranks. However, the partitioning of this variation within the 6 samplings of each quadrat indicates that 65-90% is due to variation in vegetation composition between quadrats and only 10-30% to observer and other sources of variation within quadrats.

Also from the 6 repeated observations it was possible to get a reliable estimate of the true rank order within each quadrat against which to compare the individual observations and thus get more information on observer variation at different rank values (Table 2b). The observer variation in repeated samplings shows the least standard deviation for the lowest ranked (most abundant species) and reaches a maximum standard deviation of about two rank positions for species ranked 6th to 8th. Further subdivision of this variation shows a highly significant effect (Bartlett's test) of the experienced (I & II) versus inexperienced observer (III) and of difference between vegetation types ('contrasting' = a few very dominant species and many subordinate species, versus 'similar' = vegetation where species were of more similar abundance).

There are few comparisons of observer variability, as distinct from quadrat variability, for other visual sampling methods. For cover estimates Smith (1944) showed that observer and other sources accounted for 15% of variation in 100 $ft^2$ quadrats in range land, to 6% in 200 $m^2$ in woodland. Similarly Sykes *et al*. (1983) showed that the 90% confidence interval in woodland was ±10-20% in cover estimates.

## Individual Species Analysis

*Nature of data and statistical techniques*

The data produced for individual species is both relative and ordinal. It is relative in that the rankings are only in relation to other species present. It is ordinal in that the values can only be allocated to one of a few mutually exclusive categories, but in which there is a definite relationship in magnitude between those categories. Some may share the lowest classes of 'only present' or 'absent'. However the data could also be downgraded to nominal by retaining classes but with no implied fixed relationship between classes.

The Lehmann rank contingency tables (Lehmann and D' Abrera, 1975) and the regression model for ordinal data (McCullagh, 1980), are two analysis methods which have been specifically developed for rank data of a limited number of classes. Other 'rank' analysis techniques in many older texts strictly imply a continuous scale though the adjustments for tied ranks give an approximation for data with a limited number of rank classes.

The name of 'generalized linear models' has been given to the general analysis of data in which there is subdivision into a dependent variable and associated independent covariates, and where the dependent variable is quantitative, ordinal or nominal, with fixed effect covariates and a single error term (Aickin, 1983; McCullagh and Nelder, 1983; Agresti, 1986). While the analysis options are not yet as great or as readily available as for analysis of variance or multiple regression with quantitative dependent variables, the consideration of the generalized linear model including development of techniques for ordinal data (for example, Agresti, 1984; Stanish, 1986) is a rapidly developing field and some of the approaches used in this paper are already somewhat dated.

*Summary parameters for individual species*

Ordinal data are less easy to summarize than quantitative data, though the frequency distribution in rank classes is the most basic description. Corresponding to the mean and standard deviation of quantitative data, the comparable parameters for ordinal data are the median (the central value in an ordered sequence) and the quartile range (the 25% either side of the median; Table 3). However, it is likely that the mean of the rank values for a species will be the parameter commonly used.

*Association between species*

The association between species is largely dependent on the quadrat size used. There has only recently been development of a correlation coefficient for rank data with a limited number of classes but it is not widely available (Stanish, 1986). The older Spearman and Kendall rank correlation coefficients which appear in many texts can be adjusted to allow for data that have a limited number of rank classes or ties (Siegel, 1956). The effect of these modifications is to reduce the size of the correlation coefficient. The example using the Spearman correlation coefficient shows negative association between all the species pairs, whether all data are used (Table 4, right top hand sector) or only data in which both species were common (rank less than 5; Table 4, bottom left hand sector). This is partly because the example data used common species.

The above discussion considered the species in pairs. When several species have to be considered simultaneously then principal component analysis is appropriate. This shows the degree of association between species by their proximity on a diagram. Figure 1 gives the principal component analysis using Spearman rank correlations. The wide spacing in the diagram indicates negative association between these four species.

Table 3: *Example of summary parameters for individual variables Frequency distribution of ranks (%) (n = 2358)*

| | Species | | | |
|---|---|---|---|---|
| Parameter | brown-top | hawk-weed | alsike clover | white clover |
| Ordinal or rank class | | | | |
| 1 | 19 | 25 | 28 | 20 |
| 2 | 21 | 17 | 24 | 14 |
| 3 | 16 | 11 | 13 | 9 |
| 4 | 11 | 9 | 8 | 5 |
| 5 | 7 | 7 | 4 | 4 |
| 6 = present | 10 | 7 | 5 | 7 |
| 7 = absent | 16 | 24 | 18 | 41 |
| Median rank | 3 | 3 | 2 | 5 |
| Quartile range | 4 | 4 | 4 | 5 |
| Mean rank | 3.6 | 3.7 | 3.2 | 4.4 |

Table 4: *Association between pairs of species using the Spearman correlation coefficient (n = 884). The top right-hand sector of the table is for all data and the bottom left for data only in which both species are present.*

| | browntop | hawkweed | alsike clover | white clover |
|---|---|---|---|---|
| browntop | - | -0.16 | -0.13 | -0.14 |
| hawkweed | -0.36 | - | -0.09 | -0.29 |
| alsike clover | -0.31 | -0.31 | - | -0.24 |
| white clover | -0.25 | -0.27 | -0.22 | - |

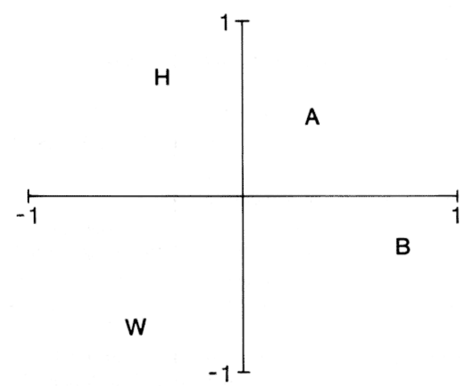


Figure 1: *Factor analysis of association between the four species. A = alsike clover, W = white clover, H = hawkweed. B = browntop.*

*Rank changes relative to qualitative or treatment variables*

In most investigations there are questions as to how the abundance of a particular species changes in response to some treatment or other variable. These other variables may be either designed treatments, such as differential grazing, or some variable relating to the sampling points such as soil type or rainfall zone. In this section are examples where these independent variables are categorical (nominal or qualitative) rather than quantitative.

Table 5 shows an example giving the mean rank, and proportion of quadrats ranked first or second, of brown top in the eight development X season of grazing treatments in the second year.

Six different analysis methods were applied to this set of data, all giving similar conclusions for practical purposes (Table 5b). All methods indicated significant difference in mean rank between 3.1 and 6.0 for the 8 treatments, and between the four season means. The significances of the development means, or the spring

Table 5: *Comparison of browntop ranking in the second year of the study in relation to qualitative or treatment variables, together with associated statistical tests, assuming either nominal, ordinal or semi-quantitative species data (n = 371). [a]Only one treatment comparison at a time, [b]Simultaneous comparison of two or more treatments, [c]After adjustment for heterogeneity.*

a) Treatment effects - as mean rank of browntop.

| Development | Spring | Summer/Autumn | Winter | Rotational | Mean |
|---|---|---|---|---|---|
| Strip | 4.2 | 4.56 | 6.0 | 3.1 | 4.5 |
| General | 3.2 | 4.3 | 4.66 | 3.7 | 4.0 |
| Mean | 3.9 | 4.5 | 5.6 | 3.3 | |

b) Treatment effects - as % quadrats in which browntop ranked 1st or 2nd.

| Development | Spring | Summer/Autumn | Winter | Rotational | Mean |
|---|---|---|---|---|---|
| Strip | 16 | 9 | 8 | 44 | 19 |
| General | 25 | 7 | 7 | 34 | 18 |
| Mean | 18 | 9 | 8 | 41 | |

c) Statistical tests (probability values for selected comparisons).

| Assumed nature of data and method of analysis | Season | Development | Interaction | Spring vs summer | 4.56 vs 4.66 |
|---|---|---|---|---|---|
| i) Rank or ordinal | | | | | |
| Lehmann contingency table[a] | 0.0001 | 0.006 | | 0.011 | 0.79 |
| Ordinal regression[b] | 0.0001 | 0.001 | 0.001 | 0.036 | 0.90 |
| Ordinal regression[c] | (0.001) | (0.030) | (0.001) | | |
| Median test non parametric[a] | 0.0001 | 0.006 | | 0.012 | 0.79 |
| Wilcoxon & Kruskal-Wallis[a] | 0.0001 | 0.006 | | 0.012 | 0.79 |
| ii) Nominal | | | | | |
| Log-linear contingency table[b] | 0.0001 | 0.03 | 0.30 | 0.17 | 0.74 |
| iii) Quantitative | | | | | |
| ANOVA | 0.0001 | 0.69 | 0.09 | 0.05 | 0.80 |

versus summer means, were marginal, and no method could discriminate between the two individual treatment means of 4.56 and 4.66 with the 93 quadrats involved.

The Lehmann contingency table and the ordinal regression analysis and its associated analysis of deviance were regarded as the most appropriate because of their retention of the species rank data in their original ordinal form. But these two methods are not available in many computer statistical packages. The median and the Kruskal-Wallis analysis re-rank the given data from highest to lowest value, the log linear contingency table downgrades it to several nominal classes, while the ANOVA requires treating the data as if it were quantitative.

There are a number of differences between treating the dependent variable data as ordinal compared with treating it as quantitative.

Firstly each of the original data observations in the ordinal regression approach only contibutes to a cell frequency and hence to the frequency distribution across one treatment combination. With quantitative data each observation makes a direct contribution. This difference influences considerations like 'degrees of freedom' in the various statistical tests.

Secondly the dependent variable in the ordinal regression is the proportion of observations or quadrats of a particular type, whereas with quantitative data the value of the dependent variable is directly estimated. Hence the second method of presenting summary data (Table 5b) is probaby more appropriate if ordinal regression analysis is being used, though it does require a decision on what combination of rank classes to use in presentation.

Thirdly, the ordinal regression method using analysis of deviance does not assume or use a balanced design in relation to treatment variables as in many analysis of variance methods. Hence it is more similar to ANOVA for unbalanced designs, where the significance depends on the order in which variables are considered, even if in practice it is a factorial complete design. Also it is recommended that subdivision of 'source of variation' and significance testing is conservative; for example, a variable's significance is tested assuming it is the last to be included in an analysis model.

Fourthly, at present ordinal regression uses a fixed model with all effects tested against the residual deviance. This presently precludes the correct analysis of split-plot experimental designs.

Finally the results as presented assume that the original sample data were independent and with constant variance. In most vegetation rank data there are likely to have been multiple quadrat samples within each treatment or other subdivision units, and hence heterogeneity of variance is also likely. This can be adjusted for in the ordinal regression approach and the value in brackets in Table 5 shows that this reduces the significance levels.

The other two methods given for rank data are similar to one way analysis of variance (Siegel, 1956). The median test is a general method and compares the median or the rank scores of a species in the different treatment levels. The Wilcoxon and Kruskal-Wallis one way ANOVA test looks further at the distribution of the ranks and tests whether higher ranks are preferentially distributed within some treatment levels.

The present species rank data, when re-ranked for the Wilcoxon or Kruskal- Wallis analysis will result in ties. Adjustment for these ties will increase the significance of the results, so that an analysis without correction will be 'conservative'.

Except for the ordinal regression the other three ordinal methods are strictly only applicable to one treatment comparison at a time, but by repeated application to the same data it is possible to get some of the features of multifactor analysis.

Another general approach is to downgrade the species data to nominal but to retain the differentiation of dependent and independent variables and to use the more recent methods developed for such functional models of categorical responses (Grizzle *et al.,* 1971; McCullagh and Nelder, 1983).

Besides the tests of significance for effects of independent variables, ordinal regression analysis allows estimation of proportions in different situations like, for example, "16% of the quadrats in the strip development treatments had browntop ranked first or second, as compared to 25% in the general developed treatments", or "the least significant difference was 13070 between the means for the seaons for quadrats in which browntop was ranked first or second".

*Changes relative to a quantitative variable*

The example in Table 6 shows an analysis of the change in ranking with time in which years have been treated as a quantitative continuous variable. The test of significance showed there was a difference between years, a significant linear trend for both species, and an additional significant quadratic effect for hawkweed. The fitted ordinal regression equation (Table 6) is superficially more complex than the more familiar multiple regression for quantitative data, but otherwise has many similarities. Note that the dependent variable is proportions of observations or quadrats exceeding a particular rank value for a particular combination of variables.

The distinction between qualitative and quantitative independent variables is somewhat artificial, as it is in normal regression, in that in reality all variables are treated as quantitative with qualitative variables treated as having two values only (0, 1 or - 1, 1).

Table 6: *Ordinal regression of species rank data on a quantitative variable (changes in hawkweed and white clover with years, see Fig.* 2).

| | Probability level | |
|---|---|---|
| | Hawkeed | White Clover |
| Tests of significance | | |
| Years as 6 levels | 0.0001 | 0.0001 |
| Years as linear trend | 0.0001 | 0.001 |
| Years as quadratic | 0.001 | 0.2 |

**Fitted model**

% quadrats more abundant than specified rank
= 100 exp (Z(l + exp Z))

Model coefficients for hawkweed

$Z = K - 1.85 + 0.36 \text{ year} - 0.21 \text{ year}^2$
± 0.12 ±0.12

where K = 1.85 (for rank 1); or - 0.68 (for ranks 1 + 2); or - 0.08 (for ranks 1 + 2 + 3).

% deviance explained 83 %

Model coefficients for white clover

$Z = K - 0.46 - 0.22 \text{ year} + 0.08 \text{ year}^2$
± 0.11 ± 0.02

where K = -1.44 (rank 1); or -0.63 (ranks 1 + 2); or -0.18 (ranks 1 + 2 + 3).

% deviance explained 27%

These particular analyses were done a number of years ago using a special purpose program PLUM. Currently they would be done using the generalised linear model within the GENSTAT package, or with PROC CATMOD in the SAS package after determining cumulative frequencies.

## Vegetation Analysis

*Vegetation types and grouping of observations*

Vegetation descriptions commonly include consideration of both species content and their relative dominance or abundance, usually by reference to the more abundant species, for example, "a fescue tussock/hawkweed depleted tussock grassland". Such an approach is used in several systems of phytosociology, such as the Braun Blanquet and Zurich-Montpellier methods. This approach can be more formalised if species are listed in their rank order. Then a vegetation type can be described as a species A/species B/species C/species D etc. community. If the ordering is according to rank abundance then the listing can differentiate between vegetation types formed from different combinations of the same species; a vegetation type of species order A/B/C is different from a vegetation type of species order C/B/A. As the relative abundance of species decreases with rank, the number of species used to define a vegetation type can be terminated as appropriate.

In practice one has to apply this concept to grouping particular sample data. The data produced by the proposed ranking method can be hierarchically sorted into groups of like rank combination - firstly according to the first ranked species in quadrats, then within those for the second ranked species, within those for the third ranked species, etc. This is a hierarchical divisive method of clustering. The potential number of such vegetation types is large. If N is the number of species present in the total sampling area, and R the number ranked at each point, then the potential number of vegetation types defined by the rank order of species combinations is N!/(N-R)!. For example with 3 species and 2 rank levels there are 6 vegetation types (A/B/C, A/C/B, B/A/C, B/C/A, C/A/B, and C/B/A); and for 10 species and 4 ranks there are 5040.

Even though the actual combinations that do occur may only be a fraction of the potential number, they are still likely to be more than can be used. The first reduction can be according to frequency, and a lower limit of between 1% and 5% of the total

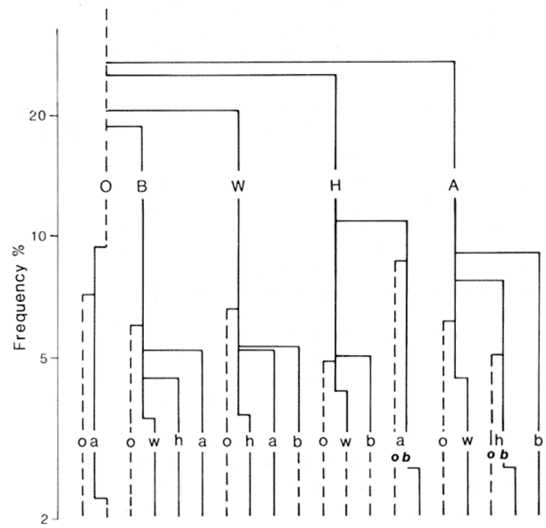


Figure 2: *Hierarchical subdivision of quadrats into vegetation types based on frequency of occurrence of rank orders of different species combinations. B = browntop, H = hawkweed, A = alsike clover, W = white clover, and 0 = 'others', First ranked species in capitals, second ranked species in lower case and third ranked species in italics.*

samples is suggested. There may need to be a compromise in accepting types based on first or second ranked species versus further subdivision of frequent types using lower ranked species.

The example used had 2358 sampling points and considered five species (brown top, hawkweed, alsike clover, white clover, and 'others'). Six vegetation types exceeded 10% of sample points using only the first and second ranked species, 11 types exceeded 5% using three ranks, while 20 types exceeded 2% (Fig. 2). Twelve vegetation types were used in the example solely to include a group with a three species identification (Table 7). The table also gives the proportion of these in the second and sixth sampling years together with some projections to be discussed in the next section.

Four features of the grouping procedure should be noted. Firstly, within a particular set of data the number of species needed to characterise the different vegetation types may vary. Secondly, the classification of types must be mutually exclusive. Thirdly, the category of species called 'others' includes not only all the species not considered independently, but often

Table 7: *Vegetation types recognised from rank order of species, their percentage contributions in the second and sixth years, and projected contributions in the tenth year and at equilibrium.*

| Code | Vegetation Description | Year | | | Projected |
|---|---|---|---|---|---|
| | First rank/Second/Third | 2 | 6 | 10 | Final |
| B/A | browntop/ alsike | 5 | 11 | 7 | 7 |
| B/O | browntop/ 'others' | 11 | 3 | 13 | 13 |
| H/A | haw kweed/ alsike | 1 | 1 | 7 | 7 |
| H/O | hawkweed/'others' | 1 | 0 | 10 | 10 |
| A/B | alsike/browntop | 1 | 35 | 10 | 10 |
| A/H/B | alsike/hawkweed/browntop | 1 | 1 | 3 | 3 |
| A/H/O | alsike/hawkweed/'others' | 0 | 2 | 5 | 5 |
| A/O | alsike/'others' | 10 | 24 | 12 | 12 |
| W/B | white clover/browntop | 17 | 3 | 7 | 7 |
| W/A | white clover / alsike | 14 | 10 | 7 | 7 |
| W/O | white clover/'others' | 24 | 6 | 12 | 12 |
| O | 'others' | 15 | 4 | 8 | 8 |

also a named species which itself is not the basis for a subdivision at a particular rank level. A further consequence of this is that the subdivision is not strictly hierarchical; that is, the initial frequency of classes including 'others' may change as further subdivisions are made. Fourthly, if the study is to consider time trends (see following section) then it is necessary that any sample points not included in the designated vegetation types must be included in the conglomerate class of 'others'.

The strength of this approach of grouping samples is that it retains the ordinal nature of the raw data and allows subdivision into vegetation type groupings, and that it is easy to allot a new sample to one of the defined groups. Its weakness is that the groupings are still categorical, and therefore it is not possible to use statistical analysis to differentiate, or test differences, between vegetations.

*Vegetation trends*

The dynamics of the vegetation can be determined from permanent quadrats subject to repeated measurements over a number of time intervals. In the past this approach has been open to the major criticism that, for use with classical statistical techniques, data from permanent quadrats lack independence. This would be a valid criticism of the regression example used in Fig. 2.

An alternative approach for use with repeated samplings is the use of transition matrices, which give the estimated probabilities of quadrats of one vegetation type changing to that of another (Austin, 1980; Enright and Ogden, 1979; Van Hulst, 1979; Debussche *et al.,* 1972). These probabilities are estimated from the relative frequency with which quadrats of each vegetation type change to other types within the selected time period. For example, of the 125 quadrats which were classified as browntop/alsike in any of the years, then 18% were still browntop/alsike the next year, as compared to the 20% which had changed to browntop/others, etc. Table 8 gives the transition matrix for data of the 12 vegetation types used in Table 1 based on changes in the yearly data from second to sixth years.

Some features of this approach should be noted. Firstly, the classes are nominal so in the present context can be either species rank classes or vegetation classes. Secondly, the matrix represents the change in one time unit - whatever that might be in the particular circumstances. This implies at least two measurements for each quadrat. Thirdly, all observations are used in developing the matrix of values. Fourthly, the vegetation classes have to be mutually exclusive but all inclusive. Finally, the reliability of an individual probability can be estimated from the binomial distribution using the number of quadrats in the margin of the species or vegetation classes (variance = np(l-p)).

A transition matrix can be looked at in a number of stages. A broad distribution of values over the matrix, as contrasted with a preponderance of both very high and very low probabilities (given as percentages), indicates a weakly structured interaction between vegetation types. The higher values in the

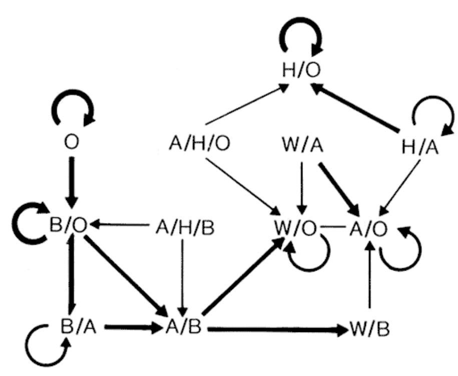


Figure 3: *Diagrammatic representation of interaction between vegetation types shown by transition matrix.*

Table 8: *Transition matrix of changes in 12 quadrat vegetation types using annual changes in combined yearly data. Probabilities recorded to nearest 1%. Values greater than 15 % are given in bold type.*

| Vegetation one year later | | Present vegetation | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | B/A | B/O | H/A | H/O | A/B | A/H/B | A/H/O | A/O | W/B | W/A | W/O | 0 |
| B/A | browntop/ alsike | **18** | **17** | 5 | 4 | 9 | 7 | 10 | 3 | 2 | 4 | 2 | 3 |
| B/O | browntop/'others' | **20** | **28** | 2 | 13 | 12 | **17** | 6 | 8 | 12 | 3 | 6 | **22** |
| H/A | haw kweed/ alsike | 1 | 1 | **15** | 11 | 0 | 4 | 10 | 6 | 11 | 10 | 9 | 6 |
| H/O | hawkweed/'others' | 2 | 4 | **26** | **23** | 0 | 9 | **18** | 6 | 12 | 8 | 7 | 11 |
| AlB | alsike/browntop | **23** | **20** | 5 | 11 | 7 | **15** | 8 | 7 | 12 | 5 | 4 | 9 |
| A/H/B | alsike/hawkweed/browntop | 1 | 2 | 3 | 5 | 0 | 7 | 4 | 2 | 5 | 2 | 4 | 1 |
| A/H/O | alsike/hawkweed/'others' | 6 | 4 | 4 | 1 | 2 | 8 | 5 | 5 | 5 | 8 | 13 | 5 |
| A/O | alsike/'others' | 7 | 11 | **16** | 7 | 3 | 13 | 5 | **18** | **15** | **20** | 13 | 9 |
| W/B | white clover/brown top | 6 | 2 | 2 | 4 | **24** | 9 | 6 | 6 | 8 | 7 | 7 | 2 |
| W/A | white cIover/alsike | 1 | 3 | 7 | 9 | 14 | 2 | 9 | 10 | 3 | 10 | 6 | 5 |
| W/O | white cIover/'others' | 3 | 3 | 9 | 23 | 7 | 15 | 18 | 11 | 15 | 18 | 5 | |
| 0 | 'others' | 12 | 5 | 6 | 5 | 6 | 2 | 4 | 11 | 4 | 8 | 11 | **22** |
| | Number of quadrats | 125 | 240 | 116 | 180 | 188 | 42 | 89 | 209 | 118 | 118 | 213 | 148 |

| Statistical tests | Probability |
|---|---|
| Dependence on previous state | <0.001 |
| Difference between 6 time periods used | <0.001 |
| Departure from uniform probabilities | <0.001 |
| Departure from constant unchanging vegetation | <0.001 |

diagonal show constancy over one year of several of the vegetation types (for example, browntop/alsike, browntop/'others', hawkweed/alsike, hawkweed/'others', alsike/hawkweed/'others', and 'others') by the relatively high percentage probabilities of remaining in the same class.

The large interactions are indicated by bold type in the table and also shown in Fig 3. These show the low level of interaction between the quadrats dominated by hawkweed, browntop, or white clover, and the interaction via alsike dominated quadrats.

The occurrence of some high probabilities well away from the diagonal shows that there can be major changes of some vegetation types within the time span of one year.

In practice the probabilities within the transition matrix are determined from frequency counts in an historical set of permanent quadrat data. However, the statistical tests of various properties of transition matrices are best applied to cases where there are non-zeros in all cells (and accordingly neither 0 or 100% in any). So in practice it is suggested that a small frequency value (say 0.2) is entered in all cells before the transition probability percentages are calculated and statistical tests made.

As the transition matrix probabilities are likely to be determined from measurement over several time periods (as in this case) one would wish to test whether the probabilities were different between the periods. Also the observed transition matrix can be compared against some hypothesised matrix. In the example the transition matrices from the individual time intervals did differ significantly indicating there was large year to year variation as well as trends. Also the combined transition matrix did differ significantly from both a hypothetical matrix of uniform values (chance occurrences) or unchanging vegetation (near 100% on diagonal).

Van Hulst (1979) describes some other statistical tests on the matrix which could be used to answer a number of biological questions. Examples are: is the change to new vegetation type dependent on the previous type, or how long had it been that type, or is it also dependent on still earlier types?

The multiplication of the frequency of vegetation types by the transition matrix allows prediction of future frequencies. Table 7 and Fig. 4 give the frequency of selected vegetation types during the measurement period (which includes both year to year variation and trends), and the extrapolation for the next 10 years (which only gives the mean trend though there would be a similar large year to year fluctuation around these). This indicates an increasing frequency of alsike/browntop types and the slower increase of browntop/'others'.

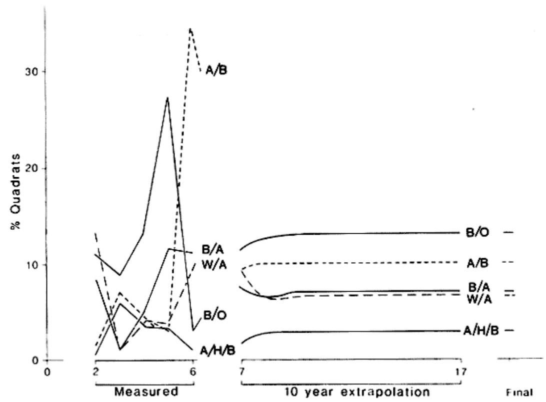


Figure 4: *Proportion of quadrats in five of the* 12 *different vegetation types during the measurement period, predicted for 10 years using the transition matrix (Table* 8), *and the predicted steady state. A = alsike, B = browntop, H = hawkweed, and O = 'others', with A/B = alsike/browntop, etc.*

One of the interesting properties of a continuous transition matrix, if all states can be achieved from another, is that the final vegetation (after an infinite time with the same conditions prevailing) is a function of the transition matrix alone and not the actual measurement period, and one can estimate what might be the proportion of quadrat types in the final vegetation. Fig. 4 shows that this distribution is not very different from that after 10 years, which in turn was not very different from that after the 6 year experimental period. The confidence intervals of these forward projections will be related via the multinomial distribution, to the number of observations in each year (380) used in forming the transition matrix.

## Discussion

The first attraction of the visual ranking technique is that it corresponds closely to the way people describe vegetation in conversation by listing species in order of importance. A further attraction is that it is generally easier to estimate the difference in abundance between species, and hence their rank, than to estimate absolute levels of each. There is also better consistency between observers, which is especially important if assessments have to be made over long intervals by different observers.

Since the only information recorded is the rank order of species at sample points, without reference to a quantitative scale, the method differs from other vegetation sampling techniques using visual estimation, or assignment to different abundance classes (Brown, 1954). The limitation of the ranking approach is that it is only a visual estimate and that it does depend on an adequate definition of the attribute being ranked.

The approach was initially used in a survey of species available to sheep in multi-species native and oversown blocks on variable terrain, and where rapid assessment had to be made at intervals of several months. More recently the method has been used for yearly assessment of permanent quadrats in long term comparisons of different oversowing and grazing treatments.

In the examples given of the statistical analysis of rank data for individual species the methods have been confined to those for ordinal classes and have demonstrated that there are appropriate and powerful techniques available.

If rank values could be given a quantitative interpretation or transformation then a greater range of techniques of statistical anlaysis would be possible. The simplest of these would be to treat the rank data as if it were quantitative continuous data. However a more justifiable approach for the transformation of rank values to a semi-quantitative scale already exists in plant and animal ecology in the relationship between number of species (diversity) and relative number or quantity of each species (abundance) within an area (see reviews by Preston, 1948; Whittaker, 1972; Pielou, 1975; Southwood, 1978). Four relationships between rank and abundance have been suggested: geometric series; logarithmic series; log normal; and broken stick series. The simplest is the geometric series which, with estimation of the ratio in abundance between any two ranked species, allows the estimation of the percentage contribution of all species to the total vegetation within the quadrat. This is why the field data included an estimation of the ratio of first to fifth ranked species. In the example the field estimate was that the fifth ranked species averaged 0.1 of the abundance of the first ranked. Using the geometric series, the relative weighting used for each rank expressed as a percentage of the total is: rank 1 = 43.8%; rank 2 = 24.6%; rank 3 = 13.8%; rank 4 = 7.8%; rank 5 = 4.4%; rank 6 (present) = 1.9% (= mean of next 5 species had they been ranked); and rank 7 (absent) = O. For example the mean for species B in Table I becomes (24.6 + 13.8 + 24.6 + 24.6)/5 = 22.2%. One advantage of this approach is that by summing the percentage contribution of the species ranked, an estimate can be obtained of the remaining vegetation not accounted for by those considered. In the example used, the sum of the contribution of the first five ranks is (43.8 + 24.6 + 13.8 + 7.8 + 4.4) = 94.4% indicating that in this example most of the information on the vegetation is obtained by ranking only five species.

Support for this transformation approach also comes from the dry weight-rank methods of botanical analysis described by Mannetje and Haydock (1963) and extensively used in Australian tropical pastures (Tothill *et al.,* 1978; Hargreaves and Jones, 1978). In that method only the first three species in a quadrat are ranked and their proportions multiplied by the empirical multipliers 70.2 for first ranked, 21.1 for second ranked, and 8.7 for third ranked to give estimates of mean percentage composition. However I believe this method works because of the relationships described in the previous paragraph and that multipliers can vary within fixed bounds (Scott, 1986).

The rank data could also be downgraded to nominal categorical data. A comparison of the different approaches applied to the same data indicates they give similar significance levels for practical purposes.

While the proposed ranking method was originally developed from the need to monitor the variability of multi-species range vegetation over a number of years in the field using the intuitive descriptions of observers, statistical techniques have been developed which are applicable to this type of data. In particular there is the relatively recent ordinal regression approach (McCullagh, 1980), in which the dependent variable takes the form of a limited number of ordinal classes, the independent variable is either qualitative or quantitative, and a fitted model is tested using standard analysis of variance and regression techniques. The present paper showed how data for a number of samples could be grouped into similar vegetation types using a hierarchical divisive method of ordinal rank value. Once the sampling points have been grouped into vegetation types then only the nominal, qualitative forms of analysis are appropriate.

Consideration of the construction of tables for frequency analysis, or of transition matrices for analysis of time trends, suggests that field data should not be condensed too quickly. The reliability of such matrices is dependent on the frequency of observations in each cell. To keep these as high as possible the data need to be retained in the smallest field observational unit or sample point. In the days when calculations were done with pencil and paper there was a strong human tendency to summarize such data plots into groups, groups into transects, transects into zones, etc., which, while retaining the mean effects, did reduce the number of units and their variability. In the age of the computer this hierarchial data reduction is not necessary and it need only be condensed immediately to the level required for each type of analysis.

The grouping of samples into vegetation types using the rank of constituent species (Table 8) lends itself to the transition matrix approach to studying time trends, as compared with the regression approach discussed earlier. The transition matrix approach gives a different view of vegetation changes with time, as compared with the regression approach. There is the attraction that the transition matrix approach actually depends on the measurement being repeated on exactly the same location as in permanent quadrats, whereas the same feature is a statistical embarrassment in the regression approach in which the different time samples must be assumed to be independent, which they are not. The transition matrix approach also gives the impression of making better use of the data from individual quadrats in that it can accommodate and usefully use the unusual but real data e.g., the loss of complete vegetation by a land-side. These outliers are again often an embarrassment in the regression approach, by increasing the variability of a trend estimate while having little effect on the mean value.

The two methods also differ in their forward predictions. In regression one can only extrapolate on the basis of the trend of the regression line and the assumed shape of the fitted line. Also fitted values at the different time intervals are generally given equal weight and it is usual to consider only polynomial relations or even linear relationships. The association between measurements in different time periods, and the placing of greater weight on the more recent measurements prior to extrapolation into the future, can be taken into account in time series regression analysis. But there seems to have been very few applications of time series analysis to vegetation data. One advantage of the regression approach is that it does not depend on measurement of equal time intervals.

Forward projection using the transition matrix approach generally indicates trends towards steady states and gives estimates as to what these might be. Also, in the few grassland vegetation data sets I have investigated, this asymptotic trend to steady state occurs in a relatively short time and implies that vegetation adjusts much more quickly to changed conditions than often thought. Asymptotic trends would not usually be considered in a regression analysis.

Another attraction of the transition matrix approach is its realistic prediction that, even with marked trends between vegetation types, most species remain present in some proportion.

In summary, the strength *of* the rank approach lies in being able to formalise the intuitive approach to vegetation description; the ease and consequent scope as a field technique; the .demonstrated approximation to more quantitative measures; the access to appropriate non-parametric and rank statistical analysis techniques; and in particular its suitability for use in the transition matrix approach to study of vegetation trends. Its weakness is its basis on visual ranking and the lack of access to the greater range of analysis procedures available for continuous quantitative measurements.

## Acknowlegements

I wish to thank Dr R. Davies, Miss E. Stevenson, Dr J. Maindonald, Miss J. Dale, and Mr A. Wallance for statistical comment, the late Dr W. Makepeace, Mr J.S. Robertson and Mr W.J. Archie, and students funded by Miss E.L Hellaby Indigenous Grassland Trust for technical assistance.

## References

Agresti, A. 1984. *Analysis of ordinal categorical data*. John Wiley, New York. 287 p.

Aickin, M.A. 1983. *Linear statistical analysis of discrete data*. John Wiley, New York. 358 p.

Austin, M.P. 1980. An explanatory analysis of grasslands dynamics. An example of lawn successions. *Vegetatio 43:* 87-94.

Brown, D. 1954. *Methods of surveying and measuring vegetation*. Commonwealth Bureau of Pasture and Field Crops, Berkshire, Bulletin 42. 223p.

Debussche, M.; Godron, M.; Lepart, J.; Romane, F. 1972. An account of use of a transition matrix. *Agronomic Ecosystems 3:* 81-92.

Enright, N.; Ogden, J. 1979. Application of transition matrix models in forest dynamics: *Araucaria* in Papua New Guinea and *Nothofagus* in New Zealand. *Australian Journal of Ecology 4:* 4-32.

Grizzle, J.E.; Starmer, CF.; Kuch, G.C. 1969. Analysis *of* categorical data by linear models. *Biometrics 25:* 489-504.

Hargreaves, J.N.G.; Kerr, J.D. 1978. BOTANAL - A comprehensive sampling and computing procedure for estimating pasture yield and composition. II. Computational package. *Tropical Agronomy Technical Memorandum 9:* 88.

Lehmann, E.L; D'Abrera, H.J.M. 1975. *Non parametrics: statistical methods based* on *rank*. Holden-Day, San Francisco. 457 p.

McCullagh, P. 1980. Regression models for ordinal data. *Journal of the Royal Statistical Society B 42:* 109-142.

McCullagh, P.; Neider, J.A. 1983. *Generalized linear models*. Chapman and Hall, London. 261 p.

Mannetje, L; Haydock, K.P. 1963. The dry-weight-rank method for the botanical analysis of pastures. *Journal of British Grassland Society 18:* 268-275.

Pielou, E.C. 1975. *Ecological diversity*. Wiley - Interscience, New York. 165 p.

Preston, F.W. 1948. The commoness and rarity of species. *Ecology 29:* 254-283.

Scott, D. 1986. The coefficients for the dry-weight rank method of botanical analysis of pasture. *Grass and Forage Science 41:* 319-321.

Scott, D.; Maunsell, LA. 1974. Diet and mineral nutrition of sheep on undeveloped and developed tussock grassland. II. Vegetation composition and availability. *New Zealand Journal of Agricultural Research 17*: 177-189.

Siegel, S. 1956. *Nonparametric statistics for the behavioural sciences*. McGraw-Hill, New York. 312 p.

Smith, A.D. 1944. A study of the reliability of range vegetation estimates. *Ecology 25:* 441-448.

Southwood, T.R.E. 1978. *Ecological methods*. 2nd edition. Chapman and Hall, London. 526 p.

Stanish, W.M. 1986. Categorical data analysis strategies using SAS software. *In:* Allen, D.M. (Editor), *Computer science and statistics: The interface,* pp. 239-256. Elsevier. 379p.

Sykes, J.M.; Horrill, A.D.; Mountford, M.D. 1983. Use *of* visual cover assessments as quantitative estimates of some British woodland taxa. *Journal of Ecology 71:* 437-450.

Tothill, J.C.; Hargreaves, J.N.G.; Jones, R.M. 1978. BOT ANAL - A comprehensive sampling and computing procedure *for* estimating pasture yield and composition. I. Field sampling. *Tropical Agronomy Technical Memorandum 8*.

Van Hulst, R. 1979. On the dynamics of vegetation; Markov chains as models of succession. *Vegetatio 40:* 3-14.

Whittaker, R.H. 1972. Evolution and measurement of species diversity. *Taxon 21:* 213-251.